마지막으로 final test를 진행한 결과입니다.

앞서 선정된 Extra Tree와 mlp, 그리고 그 두가지로 voting 을 진행했습니다

좀 더 비교가 쉽게 나타낸 막대그래프 입니다.

보이는 바와 같이, 저희는 inference time과 Accuracy를 모두 고려해, i은 낮고, A는 높은 mlp를 모델로 선정했습니다.

왼쪽은 Extra tree, 오른쪽은 mlp의 confusion matrix를 한 결과입니다.

(1, /) (4, +) (9, 7)을 잘못 라벨링한 개수가 mlp가 더 적고 - 와 =은 개수가 똑같은 결과를 볼 수 있습니다. 이 결과 역시 선별된 두 모델중 mlp가 더 Accuracy가 높다는걸 알 수 있습니다.

앞에서도 다 언급되었지만 마지막으로 데이터를 어떻게 가공하려고 했고, 마주했던 문제점이나 한계점을 얘기하면서 끝내겠습니다.

저희는 치우친 데이터도 잘 학습할 수 있으면 하는 생각에 일정 비율로 상하좌우로 shifted data를 추가해 이용하려고 했지만, 360도로 100개 정도 다른 위치에 존재하는 데이터가 아니라면 의미가 없을 거라는 피드백을 받아서 파이프라인에 중앙화 과정을 넣는 것이 효과적이었고, 최종 데이터에 중앙화 과정이 잘 이루어지는지 shifted data를 추가하는 방식으로 진행했습니다.

1 / 는 기울기 각도 차이라고 여겨 고려하지 않았고, 테두리 역시 마이너스 같은 경우에는 =으로 잘못 인식돼 제거했습니다.

다양한 노이즈가 들어와도 좋은 성능을 내는 모델을 만드는 게 목적이었고, 위 사진 속 모든 노이즈를 정확히 세세하게 제거하려면 영상처리의 범주여서 한계가 있었습니다. 그래서 데이터셋 일부에 0과 1 사이의 랜덤 값을 더해 노이즈 데이터를 생성하고, 파이프라인을 통해 노이즈를 제거하는 함수를 이용해서 진행하여 해결했습니다.